Does mutual knowledge of preferences lead to more Nash equilibrium play? Experimental evidence

Christoph Brunner[†] and T. Florian Kauffeldt[‡,⋆] and Hannes Rau[†]

Abstract

Nash equilibrium often does not seem to accurately predict behavior. In experimental game theory, it is usually assumed that the monetary payoffs in the game represent subjects' utilities. However, subjects may actually play a very different game. In this case, mutual knowledge of preferences may not be satisfied. In our experiment, we first elicit subjects' preferences over the monetary payoffs for all players. This allows us to identify equilibria in the games that subjects actually are playing (the preference games). We then examine whether revealing other subjects' preferences leads to more equilibrium play and find that this information indeed has a significant effect. Furthermore, it turns out that subjects are more likely to play maxmin and maxmax strategies than Nash equilibrium strategies.

Keywords: Behavioral Game Theory, Epistemic Game Theory, Nash Equilibrium, Incomplete Information Games, Strategic Ambiguity

JEL classifications: C91, C72

Acknowledgements. We thank Gary Charness, Peter Dürsch, Jürgen Eichberger, Paul J. Healy, Jörg Oechssler, Christoph Vanberg, the seminar participants at University of Exeter and University of Heidelberg, and the conference participants at ESA European Meeting 2014, HeiKaMax Workshop 2014, ESA World Meeting 2017, and JE on Ambiguity and Strategic Interactions 2019 for very helpful comments.

[†] University of Heidelberg, Departement of Economics, Bergheimer Strasse 58, 69115 Heidelberg, Germany.

[‡] Heilbronn University, Departement of Economics, Max-Planck-Strasse 39, 74081 Heilbronn, Germany. ⋆ Corresponding author. Email: mail@florian-kauffeldt.com

1. Introduction

People frequently exhibit behavior that seems to be inconsistent with standard Nash equilibrium behavior such as cooperation in one-shot social dilemma situations. Nash equilibria in experiments are usually determined using the monetary payoffs in the game. However, if subjects do not only care about their *own* monetary payoffs, they may actually play a very different type of game. For example, instead of facing a social dilemma game, they might actually play a coordination game (see Example 1 on page 5 for more details). This reasoning is not new and has already been discussed in the literature. For example, Weibull (2004, p. 86) writes "(...) *preferences constitute an integral part of the very definition of a game.*" Particularly in situations where monetary payoffs do not fully represent players' utilities over the outcomes of the game, it is not clear whether there is common knowledge about those preferences. We study the impact of this aspect on equilibrium play.

Common (or, at least, mutual) knowledge about preferences is a core assumption in game theory. In the words of Polak (1999, p. 673):

> "*In games of complete information, common knowledge of payoffs is usually taken to be implicit. Indeed, this is often taken to be the definition of complete information.*"[2]

Mutual or even common knowledge about preferences is not only assumed in traditional game theory, but also often in behavioral game theory. For example, most level-k models assume that payoffs are mutually known and that agents form beliefs about other agents' play based on this information (see, e.g., Costa-Gomes et al., 2001). Our findings might also be relevant to assess under which circumstances these concepts can safely be applied.

Despite the ubiquity of the (implicit) assumption of mutually or commonly known preferences, there is little empirical evidence about the degree to which it affects the reliability of the Nash prediction. However, previous experimental research suggests that it should not be taken for granted. For example, Healy (2011) finds that subjects fail to accurately predict other subjects' preferences over possible outcomes in normal-form 2×2 games. The

[2] This means that "complete information" cannot be part of the rules of the game (*the game-form*) because it involves assumptions about knowledge of individual preferences.

purpose of the experiment reported in this paper is to test whether mutual knowledge of preferences is important for the Nash prediction.[3]

Our results can be summarized as follows: (1) subjects are indeed significantly more likely to play a Nash equilibrium strategy when they are informed about their opponents' preferences over the possible outcomes of the game. When preferences are not mutually known, the frequency of equilibrium play is rather low. (2) A strategy is more likely to be played when it cannot lead to the lowest payoffs (*maxmin strategy*) or when it can lead to the highest one (*maxmax strategy*). Furthermore, maxmin and maxmax strategies predict behavior better than Nash equilibrium strategies, especially when preferences are not mutually known.

Result (1) shows that subjects not only fail to accurately predict other players' preferences, the lack of such information also significantly affects their behavior. Whenever it is unlikely that players know each other's preferences and some players have no strictly dominant strategy, it might therefore be advisable to use a more general equilibrium concept. Following Polak (1999), we may view a situation where preferences are not mutually known as a game with incomplete information. Such a situation could be modeled as a Bayesian game (Harsanyi, 1967-1968).[4]

Result (2) suggests that many subjects rely on reasoning other than Nash reasoning, especially in the baseline treatment. The reason may be that subjects face two different sources of uncertainty: in the baseline treatment, they are uncertain about the other player's preferences *and* about the other player's rationality. In the info treatment, subjects are informed about the other player's preferences but they still may not be sure about his strategy choice. When facing strategic ambiguity, maxmin and/or maxmax strategies can be a best response to it (see Eichberger and Kelsey, 2014).

[3] Notice that we examine one-shot dominance-solvable 2 x 2 games. In the context of repeated games, it has been studied whether subjects are able to learn the preferences of their opponents (see, e.g., Oechssler and Schipper, 2003). However, it is not the purpose of our experiment to examine learning in games. Repeated games require repeated interaction and are much more complex to analyze than one-shot games.

[4] Players with different preferences can be thought of as different types and it is then assumed that the prior distribution of types is commonly known.

1.1. The experiment

In stage 1 of the experiment, we elicit subjects' preferences over monetary payoff pairs (they will be referred to as "payment pairs"). The same payment pairs are then used to construct eight different 2×2 games (or more precisely eight different game-forms). In stage 2, each subject plays four out of the eight games exactly once.[5]

Our design allows us to avoid the assumption that subjects only care about their *own* monetary payments. Instead, we can use the preferences elicited in stage 1 to describe the game that our subjects play.[6]

This will be illustrated with the help of Example 1 below, which corresponds to one of the games played in the experiment.

Example 1. *Consider the prisoner's-dilemma-type game-form in Figure 1. The numbers in the matrix correspond to the amount of money paid to the players, where the first number is the row player's payment and the second number is the column player's payment.*

	L	R
U	$4,4$	$8,3$
D	$3,8$	$7,7$

Figure 1: Prisoner's-dilemma-type game-form

Depending on players' preferences over the payment pairs, various games can be induced by the game-form in Example 1. Let r be the row and c be

[5] We ran the experiment in two waves, where subjects participating in the first wave played the first four games and subjects participating in the second wave played the second four games.

[6] We maintain the assumption that preferences depend only on players' monetary payments. That is, the specific game-form, other subjects' preferences, or any other factors have no effect on subjects' ordinal ranking of payment pairs. Of course, this is to some degree a consequentialist approach and consequentialism has been criticized in the literature repeatedly. In Section 3.4, we will discuss evidence suggesting that such considerations do not play an important role in the games used in this study. We cannot completely exclude that violations of consequentialism might have caused some noise. However, we do not find any evidence that this systematically influenced our result.

the column player and denote the payment pairs by $(x_r, x_c) \in \mathbb{R}^2$. Suppose player i's ($i \in \{r, c\}$) preferences over payment pairs are represented by a function $v_i : \mathbb{R}^2 \to \mathbb{R}$. In general, the games induced by the game-form in Example 1 take the following form:

	L	R
U	$v_r(4,4), v_c(4,4)$	$v_r(8,3), v_c(8,3)$
D	$v_r(3,8), v_c(3,8)$	$v_r(7,7), v_c(7,7)$

Figure 2: Preference game of the game-form in Example 1

The preference game is only a prisoner's dilemma game if the players mainly care about their own payoffs, e.g., if $v_i(x_r, x_c) = x_i$ for $i \in \{r, c\}$. In this case, the game has only one Nash equilibrium (U, L), i.e., everyone defects. On the other hand, the game that results if players have other-regarding preferences can be very different. For instance, suppose players' preferences are represented by the following utility function $v_i(x_r, x_c) = \min\{x_r, x_c\}$ for $i \in \{r, c\}$. Then, (U, L) as well as (D, R), i.e., mutual cooperation, are Nash equilibria.

In this paper, whenever we refer to a "Nash equilibrium", we refer to a Nash equilibrium of the preference game using the preferences elicited in stage 1 of the experiment. We focus on those situations in which a unique pure Nash equilibrium exists (according to the reported preferences): one player has a strictly dominant strategy and the other player has a non-dominant unique pure Nash equilibrium strategy in the preference game.[7] Consequently, we consider situations in which subjects' opponents have a strictly dominant strategy. In the baseline treatment, the reported prefer-

[7] In our experiment, we only ask subjects to rank payment pairs ordinally. Eliciting a cardinal ranking of payment pairs would require a more complicated procedure that some subjects might fail to understand. It is not obvious that subjects can reliably assign a cardinal utility to each payment pair. As a result, we cannot compute Nash equilibria in mixed strategies for the preference games. Moreover, we will exclude the decisions of subjects who have a strictly or weakly dominant strategy in the preference game. Information about their opponent's preferences is not necessary for those subjects to compute a best response and as a result, information about the other player's preferences should not be expected to have an effect on behavior.

ences are not revealed. Hence, subjects cannot be certain that their opponents have a dominant strategy.

For example, suppose the row player in the preference game above is selfish. His pure strategy U is then strictly dominant. A column player who prefers $(4, 4)$ to $(8, 3)$ and $(7, 7)$ to $(3, 8)$ then has a unique equilibrium strategy that is not dominant: L. In treatment baseline, such a column player may not be sure whether row is selfish or not and might therefore occasionally play R rather than L.

In our second treatment (called "info"), the column player can see that row has a strictly dominant strategy and might therefore play the unique equilibrium strategy L more often. Intuitively, this logic can explain our first result that subjects are more likely to play a Nash equilibrium strategy in treatment info compared to treatment baseline. Furthermore, if a subject is uncertain about the strategy choice of his opponent, then, depending on his attitude towards uncertainty, he will try to avoid the lowest ranked payment pair (maxmin), or, to reach the highest ranked one (maxmax). This intuition explains our second result.

1.2. Related literature

The papers closest to ours are Healy (2011) and recent working papers by Wolff (2014) and Attanasi et al. (2016).

Healy examines whether the sufficient conditions for Nash equilibrium identified by Aumann and Brandenburger (1995) are satisfied when subjects play normal-form 2×2 games in the laboratory. For that purpose, subjects first choose a strategy and then report their beliefs about behavior and preferences of their opponent. Subjects' own preferences and rationality are also measured. Healy finds that there are only very few instances where all conditions are satisfied. Focusing on mutual knowledge of preferences, he finds that both players correctly predict how their opponent ordinally ranked the payment pairs in only 64% of games played. Healy concludes that "The failure of Nash equilibrium stems in a large part from the failure of subjects to agree on the game they are playing."

Since mutual knowledge of preferences is one of three conditions that are jointly sufficient for Nash equilibrium in 2×2 games (see Aumann and Brandenburger, 1995) and since the other two are also not fully satisfied in Healy's experiment, it is difficult to assess the impact of the failure of mutual knowledge of preferences on equilibrium play in isolation. By varying the information available to subjects about the opponent's preferences while holding all other factors constant, we can identify the impact of mutual knowledge on equilibrium play.

Wolff (2014) studies behavior in three-person sequential public good games. In contrast to our experiment, he does not reveal subjects' preferences over the material outcomes. Instead, he elicits subjects' best-response correspondences to the contributions of the other players. In one of his treatments, these are then revealed to all group members. This information has a much smaller effect on the frequency of equilibrium play compared to the treatment effect in our experiment.

Revealing best-response correspondences is not sufficient for subjects to be able to predict how much their opponents will contribute: Wolff measures beliefs about others' contributions to the public good and finds that subjects tend to overestimate these. As a result, they often fail to play an equilibrium strategy even though their contributions tend to be consistent with their beliefs and their own reported best-responses. As opposed to the dominance-solvable 2×2 games that we study, several iterations of alternating best responses are required in Wolff's experiment to compute the Nash equilibrium. Some subjects might not be able to do so.

Attanasi et al. (2016) also argue that when subjects have belief-dependent or other-regarding preferences, they are actually playing a game of incomplete information. In their experiment, subjects form beliefs about their opponent's type (e.g., selfish or prosocial) and choose their strategy based on these beliefs. Attanasi et al. then test whether revealing information about opponent's preferences and beliefs changes behavior in a Mini Trust Game. They find that first movers are more likely to transfer the money when they face a non-selfish trustee ("guilt-averse" trustee) and vice versa. The Mini Trust Game can be considered as a 2×2 coordination game with two pure equilibria (trust, share) and (not trust, not share). Subjects clearly coordinate better on one of these two equilibria when belief-dependent preferences are disclosed. While this result points in a similar direction as our results, Attanasi et al. do not systematically test the impact of mutual knowledge of preferences on the Nash prediction. In particular, the second movers can observe the decisions of the first movers. Therefore, the preferences of the first movers are not relevant for their strategy choices.

This paper is organized as follows. The next section describes the experimental design. We then present our results, and conclude in Section 4. The appendix provides additional information about the experiment.

2. Experimental design

Our experiment consists of two treatments (called "baseline" and "info") with two stages each. In the first stage of both treatments, we elicit subjects' preferences over eight different payment pairs. These payment pairs are then used to construct eight different 2×2 games. In stage 2, each subject plays four out of the eight games exactly once. We ran two waves of experiments. Subjects played Game 1 to 4 in wave 1 and Game 5 to 8 in wave 2.[8] In treatment "info", subjects can see their opponent's ordinal ranking of the four payment pairs used in the current game, whereas in treatment "baseline", this information is not disclosed.

2.1. Stage 1 of the experiment

Stage 1 is identical in both treatments. Subjects are asked to create an ordinal ranking over the following set X_{row} of eight payment pairs (x_r, x_c):

$$X_{row} = \{(8,3),\ (7,7),\ (5,8),\ (4,4),\ (6,2),\ (3,8),\ (3,3),\ (2,2)\} \qquad (1)$$

The first number, x_r, corresponds to the amount of money (in Euros) paid to the decision-maker in the role of a row player. The second number, x_c, is paid to some other subject in the role of a column player (the "recipient").[9] Subjects are informed that they will not interact with the recipient in any other way in either stage of the experiment.

The order in which the payment pairs appear on the screen was randomly determined beforehand and remains constant in all sessions. Subjects rank the payment pairs by assigning a number between one and eight to each pair, where lower numbers indicate a higher preference. The same number can be assigned to multiple payment pairs, thus allowing for indifference.

In treatment info, subjects are told that their rankings would be disclosed to other participants at a later stage of the experiment.[10] In treatment

[8]The games are described in detail in section 2.2.

[9]Subjects who were assigned the role of a column player ranked the same payment pairs but the first number corresponds to the other player's payoff. Rewriting X_{row} for column players such that the first number corresponds to the column player's payment and the second to the row player's, we obtain $X_{column} = \{(8,3),\ (7,7),\ (8,5),\ (4,4),\ (2,6),\ (3,8),\ (3,3),\ (2,2)\}$.

[10]We will discuss the possibility that subjects might strategically misrepresent their preferences in the results section.

baseline, we made it clear that this information would not be revealed. We will explain at the end of this section how the elicitation of preferences was incentivized. After subjects confirm their ranking, they proceed to stage 2, in which they play four one-shot 2×2 games.

2.2. Stage 2 of the experiment

We ran two waves of experiments. In the first wave, subjects played the games in Figure 3 (all numbers are payments in Euro). In the second wave, they played the games in Figure 4. All games were constructed using the same eight payment pairs, see set X_{row} defined in (1).

We made sure that the games exhibit some diversity with respect to the number of pure strategy Nash equilibria under the assumption that subjects are selfish payment maximizers. The eight games were selected on the basis of two key criteria that seem to play an important role in the context of our study:

(i) # players, who have a strictly dominant strategy (0, 1 or 2) and

(ii) # pure Nash equilibria (0, 1 or 2).

Both criteria were determined for the case where preferences correspond to monetary payoffs (i.e., on the basis of the game-forms). 2×2 games can be grouped into 6 categories based on these two criteria (some combinations are not possible, e.g., 2 players with strictly dominant strategies and 2 Nash equilibria). Games with more than 2 pure equilibria are unlikely to offer valuable insights for our analysis because they are not expected to generate many relevant observations. We first run wave 1 of the experiment, then we selected the games of wave 2 so that we have at least one game of each of the 6 categories. Furthermore, we wanted to cover most of the 2×2 games that are frequently used in experimental economics (e.g., Prisoners' Dilemma, Matching Pennies, and Battle of Sexes).

Game 1

	L	R
U	$4,4$	$8,3$
D	$3,8$	$7,7$

Game 3

	L	R
U	$4,4$	$8,3$
D	$3,3$	$7,7$

Game 2

	L	R
U	$5,8$	$7,7$
D	$6,2$	$3,3$

Game 4

	L	R
U	$8,3$	$2,2$
D	$7,7$	$3,8$

Figure 3: Games in wave 1

Game 5

	L	R
U	$3,8$	$8,3$
D	$3,3$	$7,7$

Game 7

	L	R
U	$8,3$	$6,2$
D	$7,7$	$5,8$

Game 6

	L	R
U	$8,3$	$2,2$
D	$2,2$	$3,8$

Game 8

	L	R
U	$3,3$	$8,3$
D	$2,2$	$7,7$

Figure 4: Games in wave 2

2.3. Experimental procedure, treatment information and incentives

In both treatments, subjects can see how they ranked the four payment pairs of the currently played game. This information is displayed by assigning 1-4 stars to each outcome, where more stars indicate a better outcome. In treatment info, subjects are shown both their own *and* their opponent's ranking in matrix-form (see Figure 5). Just like in the payment matrix, the first entry corresponds to the subject's own ranking while the second entry reveals the opponent's ranking. In treatment baseline, subjects are shown the same rankings matrix but this matrix only contains their own rankings.

All subjects play each of the four games of their wave exactly once, each time against a different anonymous opponent. Games are played one after another. Feedback about the outcome is only provided at the end of the experiment when subjects are paid, but not while subjects still make decisions.

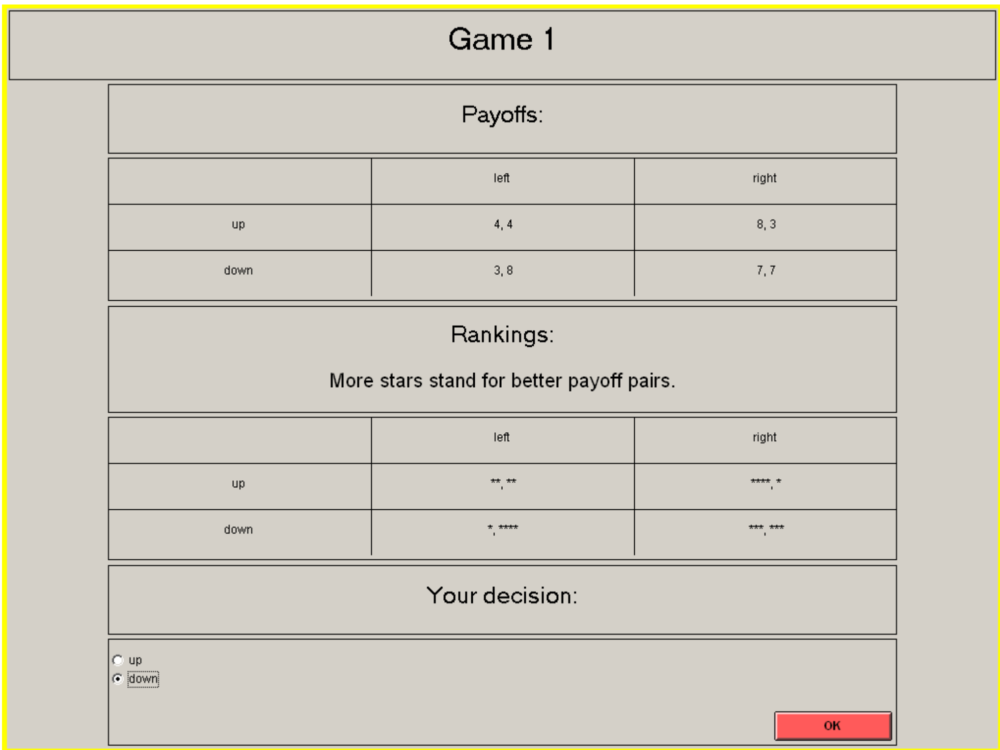


Figure 5: Information screen

In both treatments, each subject is paid for exactly one of his decisions, which is randomly selected at the end of the experiment. If a decision from stage 1 is chosen, two of the eight payment pairs from the set X_{row} are randomly selected. The row subject is then paid the first number, x_r, of the payment pair that he ranked more highly in stage 1. The second number, x_c, is paid to some other column subject. In order to avoid reciprocity considerations, we made it clear that the second number is paid to a subject with whom subjects will not interact in the second stage of the experiment. Column subjects are paid in a similar manner.

The probability that stage 1 is paid is $\frac{7}{8}$ while stage 2 is paid with a probability of $\frac{1}{8}$. These probabilities are consistent with selecting each of the $\binom{8}{2}$ possible pairs of payment pairs and each of the four decisions made in stage 2 with equal probability. Paying stage 1 with a substantially higher probability also reduces the odds that subjects might misrepresent their preferences.

This issue will be discussed in more detail in Section 3.4.

Subjects were given printed instructions and they could only participate after successfully answering several test questions. Test questions as well as the rest of the experiment were programmed using Z-Tree (Fischbacher, 2007). All sessions of the experiment were conducted at the AWI-Lab of the University of Heidelberg. Subjects from all fields of study were recruited using Orsee (Greiner, 2015). Fewer than half of the subjects were economics students. Sessions lasted about 40-50 minutes on average. The following table summarizes the number of participants per session as well as average payments:

Table 1: Summary of treatment information

Treatment	Wave	Sessions	Subjects	Average payment
baseline	1	9	97	€ 12.02
baseline	2	7	91	€ 10.54
info	1	8	95	€ 11.78
info	2	7	85	€ 11.41

Decisions made by subjects who made more than 10 mistakes when answering test questions are excluded from the data (including Table 1).[11]

3. Results

In this section, we first characterize subjects' preferences as measured in stage 1 of the experiment. We then present the main treatment effect: subjects are significantly more likely to play their unique equilibrium strategy in treatment info than in treatment baseline. This effect can be observed in 6 of the 8 games. Subsequently, we show that maxmin and maxmax strategies are more likely to be played in both treatments. We argue that it is unlikely

[11]The main treatment effect (Table 5) is still significant when these 10 subjects are included. In treatment baseline, 2 subjects made more than 10 mistakes, in treatment info, there were 8 such subjects. It is not plausible that the decisions of the excluded subjects affected other subjects' decisions since all of our games are simultaneous games and subjects were not informed about the decisions of their opponents during the experiment.

that subjects misrepresent their true preferences or that many preferences changed when subjects are shown their opponents' preferences.

3.1. Characterization of measured preferences

In stage 1 of the experiment, we elicit subjects' preferences over the payment pairs $(x_r, x_c) \in X_{row}$ defined in equation (1). Tables 2 and 3 show for each treatment the ordinal rankings reported by at least two subjects who were assigned the role of a row and column player respectively. In addition, we classify all rankings by using three categories of preferences, see below. Payment pairs that are assigned a lower number are preferred to payment pairs with a higher number.

Table 2: Preferences reported by at least two subjects who were assigned the role of a row player, both treatments. Smaller numbers are assigned to better ranked payment pairs.

(8,3)	(7,7)	(5,8)	(4,4)	(6,2)	(3,8)	(3,3)	(2,2)	n_bl	n_info	category
1	2	4	5	3	6	7	8	27	36	selfish
1	2	4	5	3	7	6	8	8	7	selfish
2	1	4	5	3	6	7	8	13	2	prosocial
1	2	4	5	3	6	6	8	6	6	selfish
2	1	3	5	4	6	7	8	4	6	prosocial
2	1	3	6	4	5	7	8	5	2	prosocial
1	2	3	5	4	6	7	8	3	2	prosocial
2	1	3	4	5	6	7	8	2	3	prosocial
3	1	2	5	6	4	7	8	1	2	prosocial
3	1	2	5	5	3	7	8	1	2	prosocial
1	2	3	5	3	6	7	8	1	1	prosocial
1	2	3	4	5	6	7	8	0	2	prosocial
3	1	2	6	5	4	7	8	1	1	prosocial
3	1	2	4	5	6	7	8	0	2	prosocial
1	1	4	5	3	6	7	8	1	1	prosocial
1	2	5	4	3	7	6	8	2	0	IA/other
1	2	4	6	3	5	7	8	1	1	prosocial

Table 3: Preferences reported by at least two subjects who were assigned the role of a column player, both treatments. Smaller numbers are assigned to better ranked payment pairs.

(8,3)	(7,7)	(8,5)	(4,4)	(2,6)	(3,8)	(3,3)	(2,2)	n_bl	n_info	category
2	3	1	4	7	5	6	8	37	31	selfish
3	2	1	4	7	5	6	8	8	6	prosocial
3	1	2	4	7	5	6	8	7	7	prosocial
3	1	2	5	6	4	7	8	5	3	prosocial
1	3	1	4	7	5	5	7	3	2	selfish
3	1	2	4	8	6	5	7	2	3	IA/other
1	3	1	4	8	6	5	7	2	2	selfish
3	2	1	5	7	4	6	8	2	2	prosocial
2	3	1	4	7	5	5	7	3	1	selfish
3	2	1	5	6	4	7	8	0	3	prosocial
1	3	1	4	7	5	6	8	1	2	selfish
4	1	2	3	6	5	7	8	1	1	prosocial
1	3	2	4	8	6	5	7	0	2	selfish
3	1	2	4	6	5	7	8	2	0	prosocial
2	3	1	4	6	5	7	8	2	0	prosocial
3	2	1	4	8	6	5	7	2	0	IA/other
3	1	2	4	8	7	5	6	1	1	IA/other

To characterize subjects' preferences, we introduce three mutually exclusive categories of social preferences: selfish preferences, prosocial preferences and inequality averse/other preferences. These categories are defined as follows:

Definition 1 (Selfish preferences)**.** A subject‘s preferences are said to be selfish, if the subject strictly prefers a payoff tuple x over a tuple y whenever his monetary payoff in x is strictly higher than in y.

Definition 2 (Prosocial preferences)**.** A subject's preferences are said to be prosocial, if the subject at least once strictly prefers a payoff tuple x to a tuple y where his monetary payoff in x is strictly lower than in y and the other player's monetary payoff in x is strictly higher than in y. In the rest of the cases, the ranking needs to be consistent with selfish preferences.

Definition 3 (Inequality averse/other preferences)**.** All preference rankings that are neither selfish nor prosocial. These mostly corresponds to some form of inequality averse preferences.

Table 4 shows the fraction of subjects whose preferences are consistent with the properties defined above.

Table 4: Measured preferences

Treatment	Selfish preferences	Prosocial preferences	IA/other preferences	Total
Pooled	48.6% (179)	39.8% (146)	11.7% (43)	368
Baseline	46.8% (88)	41.5% (78)	11.7% (22)	188
Info	50.6% (91)	37.8% (68)	11.7% (21)	180
Fisher's exact test	p=0.53	p=0.52	p=1.00	

Number of obs. for each category in brackets

The majority of reported rankings are consistent with selfish preferences, closely followed by prosocial preferences. We run a Fisher's exact test to check if there are significant differences between the treatments with respect to reported preferences. There are no significant differences between the treatments for any of the preference categories. This provides evidence that there is no (systematic) misrepresentation of preferences between treatments. Furthermore, if people misrepresented their preferences for strategic reasons and/or image concerns, we would expect to observe a lower frequency of selfish preferences in treatment info than in baseline. However, our data shows exactly the opposite.

3.2. Nash equilibrium play

Our first hypothesis is that subject behavior is more consistent with the Nash equilibrium when preferences are mutually known. We test this hypothesis by using two different subsets of our data, which are depicted in Figure 6. Notice that we use the preferences elicited in stage 1 to identify dominant and equilibrium strategies.

There are a total of 368 subjects who participated in the experiment. Since each subject played four games (=four decisions) in stage 2, we have data on 1472 individual decisions, 752 in treatment baseline and 720 in treatment info. Only one type of strategic situation is of interest for examining the effect of mutually known preferences on equilibrium play: **a unique pure Nash equilibrium where exactly one player has a strictly dominant strategy**. Therefore, we exclude those decisions where both strategies are played with strictly positive probability in some Nash equilibrium. Furthermore, we exclude those decisions where the equilibrium strategy is weakly or strictly dominant. In such a situation, the best response does not depend on the other player's strategy and therefore, it should not matter whether or not the other players' preferences are known. Thus, we focus on dominance-solvable games in which exactly one player has a strictly dominant strategy. Consequently, the other player has a non-dominant but unique equilibrium strategy. In this situation, the disclosure of preferences might help the player without a dominant strategy to figure out his unique equilibrium strategy. This leaves us with 279 relevant individual decisions, 140 in treatment baseline and 139 in treatment info. We test our main hypothesis using these 279 observations and will refer to the corresponding subset of our data as **"all subjects"**.

We also consider a subset of the set "all subjects" which no longer includes the decisions made by subjects who played a strictly dominated strategy in at least one of the four games. Either the preferences that these subjects reported in stage 1 do not reflect their true preferences or they are not rational in the sense that their choice in stage 2 is inconsistent with their reported preferences. Table 7 shows that approximately one fourth of our subjects violate strict dominance at least once. Removing the choices made by inconsistent subjects therefore further reduces the number of observations to 226 individual decisions, 115 in treatment baseline and 111 in treatment info. We will refer to this subset of our data as **"consistent subjects only".**

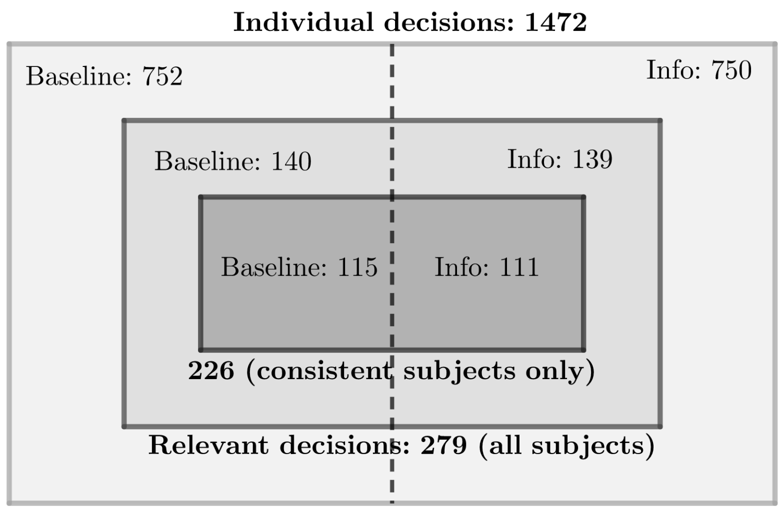
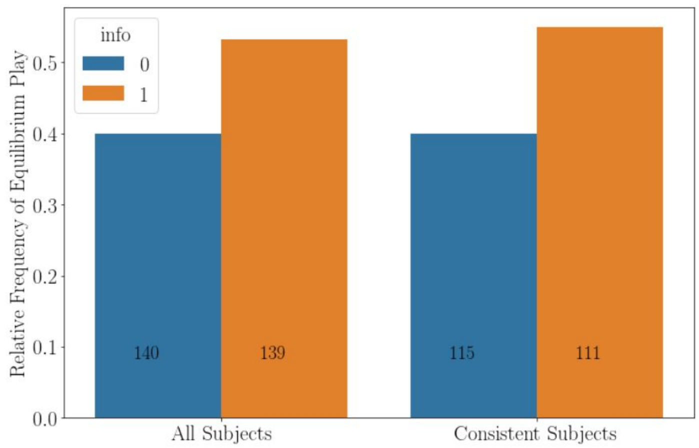


Figure 6: Relevant observations for equilibrium analysis

Figure 7 shows that subjects play an equilibrium strategy more often in treatment info than in treatment baseline. We find that the frequency of equilibrium play is 13% higher in treatment info than in baseline when considering "all" subjects and 15% higher for "consistent" subjects. Clearly, knowledge of preferences is a relevant factor affecting equilibrium play.

To test whether these differences are statistically significant, we use a linear probability model. The dependent variable "equilibrium strategy played" assumes a value of 1 if a subject plays the unique equilibrium strategy and 0 otherwise. We include an intercept as well as a dummy variable, which assumes a value of 1 if the observation is generated in treatment info and 0 otherwise. These results are shown in Table 5. The treatment effect is significant indicating that informing subjects about their opponents' preferences leads to a higher frequency of equilibrium play (p=0.037 for all subjects and p=0.036 for consistent subjects). Furthermore, the treatment effect is comparable when we only use the decisions made by consistent subjects, even though the number of observations is reduced by approximately 20%.



Bars show no. of obs. per category

Figure 7: Frequencies of played unique equilibrium strategies

As a robustness check we run a two-tailed test of proportions, where we account for multiple observations by the same person. Each person counts as one cluster.[12] The dependent variable is the frequency of equilibrium play per treatment. We run the same test for all subjects and for consistent subjects only. The null hypothesis that the distribution of the frequency of equilibrium play is the same in both treatments can be rejected regardless of which data set we use. The result from the test of proportions is in line with the result obtained from the linear probability model and confirms our main result (Result 1). In both cases, the treatment effect is significant at the 5% level.[13]

[12]We use a value for the intra cluster correlation of 0.13, which we obtained from our data. Clusters do not seem to have a strong impact on our results, because in most cases we only have one relevant decision from one subject (the average number of relevant decisions per subject is 1.3).

[13]p=0.032 for all subjects and p=0.029 for consistent subjects.

Result 1. *Subjects are more likely to play their unique Nash equilibrium strategy when preferences are mutually known.*

Table 5: Linear probability model "equilibrium strategy played", robust standard errors clustered by subject

Dependent variable: equilibrium strategy played	All Subjects	Consistent subjects only
info	0.13**	0.15**
	(0.06)	(0.07)
constant	0.40***	0.40***
	(0.04)	(0.05)
n	279	226
Clusters	212	166
Pseudo R^2	0.018	0.022

** significant at 5% level, *** significant at 1% level

Furthermore, we compute the frequency of equilibrium play for each game separately. These results are shown in Figure 8 for all subjects and in Figure 9 for consistent subjects only. The figures also include the number of relevant decisions per game. Regardless of which subset of our data we use, for most of the games, the frequency of equilibrium play is higher in treatment info than in treatment baseline.[14]

The strength of the effect varies across the eight game-forms: for instance, there is a comparatively strong effect in all games that are based on a game-form in which at least one player has a strictly dominant strategy if monetary payoffs are taken as utilities (Games 1, 3, 7, 8). By contrast, the results in games 5 and 6 seem inconsistent with the hypothesis that knowledge of preferences leads to more equilibrium play.

[14]Using a Fisher exact test, this difference is significant at the 5% level for Game 3, when we use all subjects. We have more observations for Game 3 than for any other game. In Game 3, it occurred particularly often that one subject had a strictly dominant equilibrium strategy while the other subject did not have a strictly or weakly dominant strategy. Details of these tests can be found in the appendix (tables A.9 and A.11).

At first glance, results in Game 5 appear to be surprising: there is less equilibrium play in treatment info than in baseline. This result is mainly driven by the high rate of equilibrium play in treatment baseline in this game (82.4%). Note, that our main result only allows to make statistical inferences. It does not imply that the frequency of equilibrium play necessarily has to be higher in every single game of treatment info: suppose equilibrium play is a binomially distributed random variable where the probabilities of success correspond to those observed in our experiment (p_base=0.40 and p_info=0.53). The probability to observe more equilibrium play in treatment baseline than in treatment info in a single game with an average number of subjects (n=17) is then 17.3%. As a result, the probability of observing more equilibrium play in treatment baseline in at least one out of eight games is 78.1%. In that sense, it is not surprising that the treatment effect in our experiment is negative in one out of eight games.
Moreover, Game 5 exhibits a special feature, which can also explain the observed negative treatment effect to some extent: in general, subjects are less likely to play their equilibrium strategy when it does not correspond to their maxmax strategy. In Game 5, this situation occurs substantially more often in treatment info than in baseline.[15] A comparison across all eight games reveals that this treatment difference in how often the maxmax and equilibrium strategy coincide is larger in Game 5 than in any other game (see the figure in Appendix A.2 for details).

Game 6 is a battle of the sexes game when monetary payoffs correspond to utilities. Since the preference game almost always also corresponds to a battle of the sexes game, both players have two equilibrium strategies. Since we cannot identify equilibrium play when both strategies are equilibrium strategies, few relevant observations are generated and there is no significant treatment effect due to the small number of observations.

All in all, results in individual games are in line with our hypothesis. As described in the design section, we intended to cover all common classes of 2x2 games. This general approach comes at the cost of a weaker overall effect because we also included games where we did not expect to find a strong effect.

[15] The observed difference across treatments is caused by different frequencies of prosocial players in the set of observations relevant for the analysis of Game 5.

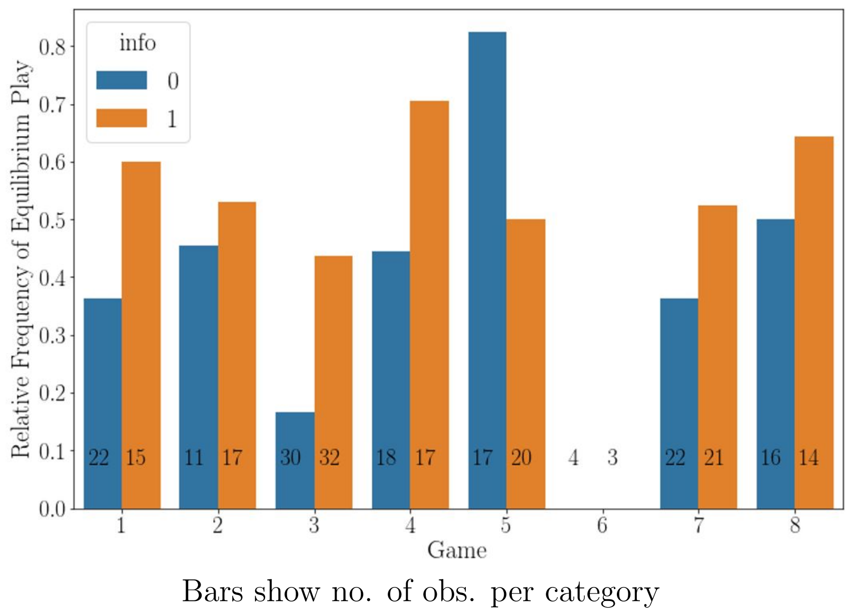


Figure 8: Frequency of equilibrium play by game, all subjects

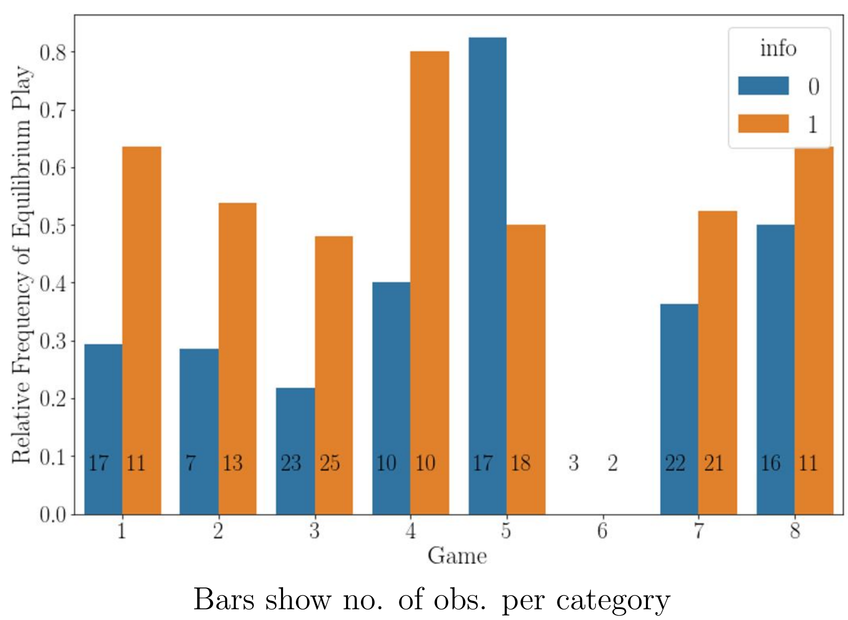


Figure 9: Frequency of equilibrium play by game, consistent subjects only

3.3. Maxmin and maxmax strategy play

As outlined in the introduction, playing a maxmin or a maxmax strategy can be a response to strategic uncertainty. Therefore, our second hypothesis is that a strategy is more likely to be played when it is a maxmin and/or a maxmax strategy. Subjects may use these strategies when they are uncertain about other players' payoff functions and/or other players' rationality. In treatment baseline, subjects face both types of uncertainty, whereas the uncertainty about other players' payoffs is removed in treatment info. Since there is some uncertainty in both treatments, we would expect a strategy to be played more often if it is a maxmin or a maxmax strategy in both treatments. Both effects are expected to be stronger in treatment baseline compared to treatment info.

We test these conjectures by running a conditional logit regression. An observation corresponds to a pure strategy. The dependent variable ("played") assumes a value of 1 if a strategy is played and 0 otherwise. Three independent variables are used to characterize each strategy: "equilibrium" indicates whether a strategy is a Nash equilibrium strategy. "maxmax" assumes a value of 1 if a strategy can lead to a most highly ranked payment pair. "maxmin" indicates whether a strategy can result in the realization of a lowest ranked payment pair (maxmin = 0 if that is the case, maxmin = 1 otherwise).

To get a more comprehensive picture about the relevant properties of the strategy selection process, we use the data from all preference games for this analysis (and not only when subjects have a unique equilibrium strategy). We also report results from the same set of data like we used for the equilibrium analysis in the appendix (table A.13). In both cases, we only use decisions made by consistent subjects. Table 6 shows that whether or not a strategy is a Nash equilibrium strategy only matters in treatment info when predicting which strategies subjects will play. In contrast, the coefficients of maxmax and maxmin are highly significant in both treatments. While the three independent variables ("equilibrium", "maxmax" and "maxmin") are correlated, all pairwise correlation coefficients are lower than 0.5. Further details on the relationship of the three independent variables can be found in the appendix.

Result 2. *In both treatments, a strategy is more likely to be played when it cannot lead to the lowest ranked payment pair and when it can lead to the highest ranked payment pair.*

In line with Result 1, the coefficient estimate for the variable "equilibrium" differs significantly among the two treatments and is only useful to predict play in treatment info but not in treatment baseline. In contrast, the highest and lowest ranked payment pair seems to attract our subjects' attention in both treatments. As expected, the according coefficient estimates are higher in treatment baseline than in treatment info. However, the difference is not significant.[16]

Table 6: Conditional logit regression "played", robust standard errors clustered by subject

Dependent variable: played	Baseline	Info
equilibrium	0.09	0.89****
	(0.21)	(0.22)
maxmax	1.61****	1.17****
	(0.18)	(0.15)
maxmin	1.39****	1.29 ****
	(0.17)	(0.15)
n	1112	1016
Clusters	139	127
Pseudo R^2	0.40	0.41

**** significant at 0.1% level

3.4. Did we manage to elicit subjects' true preferences?

To be able to clearly interpret our results, it is important to examine whether the preferences elicited in the first stage of the experiment correspond to subjects' true preferences over the outcomes of the games played in the second stage. In this section, we therefore run several tests to examine possible discrepancies between reported preferences and true preferences. We will first discuss discrepancies that might arise because subjects strategically

[16]The coefficient estimate of an interaction term of maxmin and the treatment dummy (maxmax and the treatment dummy) is not significant.

misrepresent their preferences. Discrepancies due to non-consequentialist preferences will be addressed next.

The tests that we can run using our data are not sufficiently powerful to allow us to rule out that some subjects fail to report their true preferences. However, we find no indications that reported preferences substantially deviate from true preferences. More importantly, we find no evidence suggesting that any such deviations might influence our two treatments to a different extent and thus affect our treatment effect.

Intentional misrepresentation

Subjects may have intentionally misrepresented their preferences in stage 1 for either one of two reasons:

(1) Strategic Misrepresentation: subjects in treatment info knew that their reported preferences would be disclosed in stage 2. As a result, some subjects might have tried to obtain a strategic advantage by reporting false preferences.

(2) Image Concerns: some subjects may have tried to convey a more positive image to other subjects or the experimenter by reporting false preferences (see, e.g., Hoffman et al., 1996).

In treatment baseline, it was clear to subjects that their reported preferences would not be revealed to anyone but the experimenter. If many subjects in treatment info intentionally misrepresented their preferences, we would therefore expect reported preferences to differ across the two treatments. More specifically, we would expect subjects in treatment baseline to more often report selfish preferences. That is not the case (see Table 4). In fact, subjects in treatment info are even slightly more likely to report selfish preferences (50.6% treatment info vs. 46.8% treatment baseline). Reporting selfish preferences seems inconsistent with misrepresentation either for image or for strategic reasons. Therefore, comparing the distribution of preferences across treatments suggests that that the misrepresentation of preferences due to either strategic or image concerns does not appear to play an important role in our experiment.

An additional test consists in comparing the frequency of violations of strict dominance across treatments. If subjects more often misrepresent their preferences in treatment info compared to treatment baseline, violations should occur more frequently in treatment info. That is not the case (see Table 7). The corresponding differences between the two treatments are not significant.[17]

Moreover, misrepresenting preferences in stage 1 is costly, since the first stage is much more strongly incentivized than the second stage.[18] These costs are known to subjects while the potential benefits of misrepresenting preferences are ambiguous since the specific games played in stage 2 are only revealed after preferences have been reported.

For these reasons, it seems unlikely that intentional misrepresentation of preferences plays an important role in our experiment or that such concerns could affect our treatments differently and thus explain our treatment effect.

Table 7: Violations of strict dominance

Treatment	Subjects	Games played	Games with dominant strategy	Dominated strategy played	Inconsistent subjects*
Baseline	188	752	280	23.2%	26.1%
Info	180	720	295	24.4%	29.4%

*subjects who played dominated strategy at least once

Non-consequentialist preferences

Subjects' evaluation of payoff tuples may be context-dependent. More specifically, the preferences subjects report in stage 1 of the experiment may change

(1) in the context of the game form;

(2) in the context of their opponents' reported preferences.

Game form: One possible reason why the preferences reported in stage 1 might change in the context of the game forms revealed in stage 2 could

[17] A test of proportions shows no significant differences in the average frequencies of violations of dominance per treatment (p=0.75, accounting for clusters on subject level)

[18] Recall that a decision from the first stage was paid out with probability 7/8.

be that the interests of the other player are taken into account to a larger extent in the context of a game. Since the game forms are the same in both treatments, such an adjustment of preferences should affect both treatments equally. Moreover, as discussed above, violations of strict dominance occur equally frequently in both games. That is also true when comparing the frequency of violations of strict dominance separately for each one of the eight games (results are presented in Table A.10 in the appendix). While it is possible that some subjects' preferences change in the context of the game forms presented to them in stage 2, it appears likely that such changes would affect both treatments equally. Therefore, such adjustments could explain why subjects frequently fail to play their presumed equilibrium strategy but they cannot explain the observed treatment effect.

Opponent's preferences: In many experiments, a significant share of subjects exhibit reciprocal preferences (cf. Rabin (1993) and Dufwenberg and Kirchsteiger (2004)). Observing the preferences reported by their opponents might affect how reciprocal players rank the payoff tuples of the games played in stage 2 of our experiment. Since opponents' preferences can only be observed in treatment info, such an effect could influence the measured treatment effect. To assess to what extent reciprocity or similar concerns could affect our results, we again use the frequency of violations of strict dominance as an indicator for preference adjustments. Whether reciprocal players' preferences change as a response to revealing the opponent's preferences presumably depends on the nature of the opponent's reported preferences. Therefore, we compute the frequency of dominance violations separately for each type of the opponent's reported preferences (see Table 8). Regardless of whether subjects play a selfish or a prosocial opponent, they always violate strict dominance approximately one fourth of the time. This test therefore suggests that reciprocal agents who adjust their preferences only when facing a specific type of opponent are relatively rare in our experiment.

Moreover, if such preference adjustments occurred frequently, we would have measured preferences more accurately in treatment baseline than in treatment info. This added noise in treatment info could lead to an underestimation of the frequency of equilibrium play in treatment info and thus to an underestimation of the treatment effect. Also recall that our treatment effect is larger when we only use decisions made by subjects who never violate strict dominance. These subjects are probably less prone to preference reversals, thus also suggesting that preference reversals lead to an under- rather than an overestimation of our treatment effect.

Table 8: Violations of strict dominance in treatment info

Selfish opponent	Prosocial opponent	IA/other opponent	Pooled
24.5% n=139	24.6% n=115	27.6% n=29	24.4% n=295*

*12 observations were not classifiable as there were paired with a subject excluded from the dataset

To sum up, we cannot exclude that there are some preference reversals resulting from non-consequentialist preferences. One reason is that using our data, we cannot fully distinguish between irrational behavior and non-consequentialist preferences. However, the tests discussed above provide no evidence suggesting that either non-consequentialist preferences or the intentional misrepresentation of preferences significantly affect our results, primarily because most types of preference adjustments would equally affect both treatments.

4. Conclusion

The assumptions that monetary payoffs in strategic situations represent players' utilities and that their preferences are mutually known are often not satisfied. Our experiment shows that both criteria play an important role: first, the game that subjects actually play is often very different from the one in which monetary payoffs represent utilities. Second, the treatment effect shows that mutual knowledge of preferences leads to significantly more equilibrium play. Hence, alleged violations of Nash equilibrium can be attributed, to some degree, to the violations of these two assumptions.

The main result of this study suggests that subjects fail to accurately predict other players' preferences and that this significantly affects their behavior. It is plausible that similar difficulties exist in many real-world situations as well because people often have no precise information about how other people evaluate the outcomes of an interaction. Many models that are used in behavioral game theory rely on the assumption of mutual knowledge of preferences. Since in the games we analyzed, subjects often fail to play a Nash equilibrium strategy when preferences are not mutually known, these models also might fail to accurately predict behavior whenever this assumption is not met. Therefore, when deciding what model to apply for analyzing a specific situation, the question whether or not agents can reasonably be expected to know other agents' preferences should play an important role.

References

Attanasi, G., Battigalli, P. and Nagel, R. (2016), 'Disclosure of belief-dependent preferences in a trust game', *Mimeo* .

Aumann, R. and Brandenburger, A. (1995), 'Epistemic conditions for nash equilibrium', *Econometrica* **5**, 1161–1180.

Costa-Gomes, M., Crawford, V. P. and Broseta, B. (2001), 'Cognition and behavior in normal-form games: An experimental study', *Econometrica* **69**, 1193–1235.

Dufwenberg, M. and Kirchsteiger, G. (2004), 'A theory of sequential reciprocity', *Games and Economic Behavior* **47**, 268–298.

Eichberger, J. and Kelsey, D. (2014), 'Optimism and pessimism in games', *International Economic Review* **55**, 483–505.

Fischbacher, U. (2007), 'z-tree: Zurich toolbox for ready-made economic experiments', *Experimental Economics* **10**, 171–178.

Greiner, B. (2015), 'Subject pool recruitment procedures: organizing experiments with orsee', *Journal of the Economic Science Association* **1**, 114–125.

Harsanyi, J. (1967-1968), 'Games with incomplete information played by 'bayesian' players, parts i-iii.', *Management Science* **14**, 159–182, 320–334, 486–502.

Healy, P. (2011), Epistemic foundations for the failure of nash equilibrium. Mimeo.

Hoffman, E., McCabe, K. and Smith, V. L. (1996), 'Social distance and other-regarding behavior in dictator games', *American Economic Review* **86**, 653–660.

Oechssler, J. and Schipper, B. C. (2003), 'Can you guess the game you are playing?', *Games and Economic Behavior* **43**, 137–152.

Polak, B. (1999), 'Epistemic conditions for nash equilibrium, and common knowledge of rationality', *Econometrica* **67**, 673–676.

Rabin, M. (1993), 'Incorporating fairness into game theory and economics', *American Economic Review* **83**, 1281–1302.

Weibull, J. (2004), Testing game theory, *in* S. Huck, ed., 'Advances in understanding strategic behaviour', New York: Palgrave Macmillian, pp. 85–104.

Wolff, I. (2014), When best-replies are not in equilibrium: Understanding cooperative behaviour. Mimeo.

Appendix A. (for online publication only)

Appendix A.1. Details of the robustness checks of the main result

Tables A.9 and A.11 report the results of a two-tailed Fisher exact test of the null hypothesis that the probability that a subject plays the equilibrium strategy is the same in both treatments. These tests were run separately for each of the 4 games. n_base is the number of observations in treatment baseline and n_info the number of observations in treatment info. The tests reported in Table A.9 include all subjects while those reported in Table A.11 include consistent subjects only.

Table A.9: Frequency of equilibrium play per game, all subjects

Game	n_base	n_info	p-value*
1	8/22	9/15	0.193
2	5/11	9/17	1.000
3	5/30	14/32	0.028
4	8/18	12/17	0.176
5	14/17	10/20	0.082
6	0/4	0/3	n.A.
7	8/22	11/21	0.364
8	8/16	9/14	0.484
Pooled	56/140	74/139	0.031

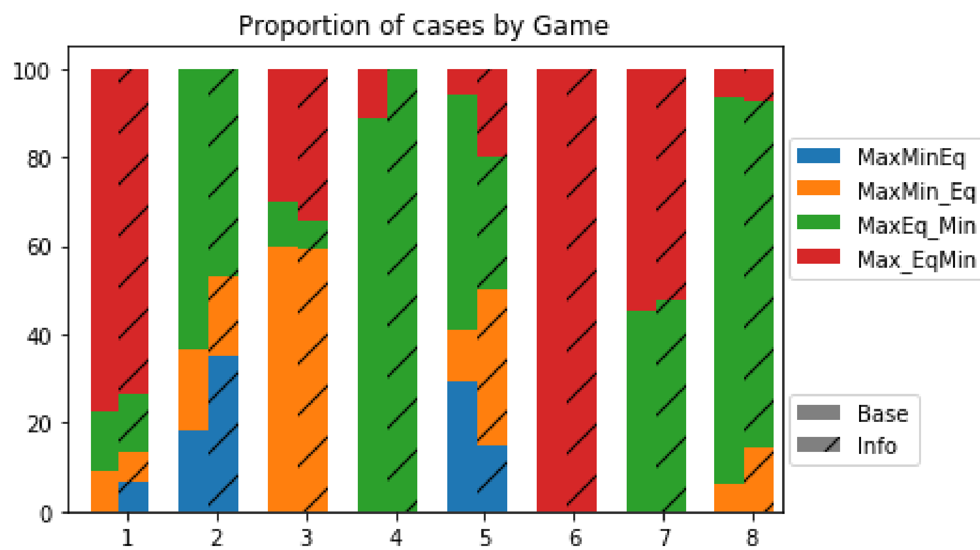
*Fisher exact test (two tailed)

Table A.10: Dominance violations per game

Game	baseline	info	p-value*
1	15/62	16/64	1.00
2	3/16	3/19	1.00
3	11/38	18/46	0.36
4	9/42	9/30	0.42
5	9/17	10/25	0.53
6	0/5	0/3	n.A.
7	10/47	11/55	1.00
8	8/53	5/53	0.56
Pooled	65/280	72/295	0.77

*Fisher exact test (two tailed)

Appendix A.2. Details on properties of the equilibrium strategy by game



This figure shows for each game how often the equilibrium strategy simultaneously is also a maxmax or a maxmin strategy. We distinguish the cases when the equilibrium strategy corresponds to both strategies (blue area), corresponds only to the maxmax strategy (green area), only to the maxmin strategy (red area) or to none of these (orange area).

Table A.11: Frequency of equilibrium play per game, consistent subjects

Game	n_base	n_info	p-value*
1	5/17	7/11	0.121
2	2/7	7/13	0.374
3	5/23	12/25	0.075
4	4/10	8/10	0.170
5	14/17	9/18	0.075
6	0/3	0/2	n.A.
7	8/22	11/21	0.364
8	8/16	7/11	0.696
Pooled	46/115	61/111	0.033

*Fisher exact test (two tailed)

Appendix A.3. Details on the conditional logit regression (table 8)

Table A.12 provides additional information for both regressions that are displayed in table 6: We check for each pure strategy available to consistent subjects whether it is a Nash equlibrium strategy ("equilibrium" = 1), whether it is the maxmax strategy ("maxmax" = 1) and whether it is the maxmin strategy ("maxmin" = 1). "n_baseline" indicates the number of pure strategies in treatment baseline, "n_info" the number of strategies in treatment info.

Table A.12: Properties of strategies available to consistent subjects, by treatment

equilibrium	maxmax	maxmin	n_baseline	n_info
0	0	0	332	322
0	1	0	236	199
1	1	1	187	189
0	0	1	182	143
0	1	1	61	56
1	1	0	56	44
1	0	1	34	37
1	0	0	24	26

Table A.13: Conditional logit regression "played" (same data as equilibrium analysis), robust standard errors clustered by subject

Dependent variable: played	Baseline	Info
equilibrium	-0.22	0.41**
	(0.19)	(0.20)
maxmax	1.07***	0.79***
	(0.30)	(0.22)
maxmin	0.93***	0.82 ***
	(0.31)	(0.22)
n	280	278
Clusters	107	105
Pseudo R^2	0.14	0.11

** significant at 5% level, *** significant at 1% level

Appendix A.4. Experiment instructions

Treatment Baseline: Instructions Part 1

1 General Information

Welcome to this experiment and thank you very much for your participation! Please switch off your mobile phone now and do not communicate with each other any more. If you have a question, raise your hand, we will come over to your seat and answer it individually. In this experiment, you can earn a substantial amount of money. The amount you earn depends on your own decisions, the decisions of the other participants and on chance. The amount of money earned will be paid out to all participants individually in cash at the end of the experiment. During the experiment, everyone makes his decisions anonymously on his own. At no point in time will your decisions be linked to your identity.

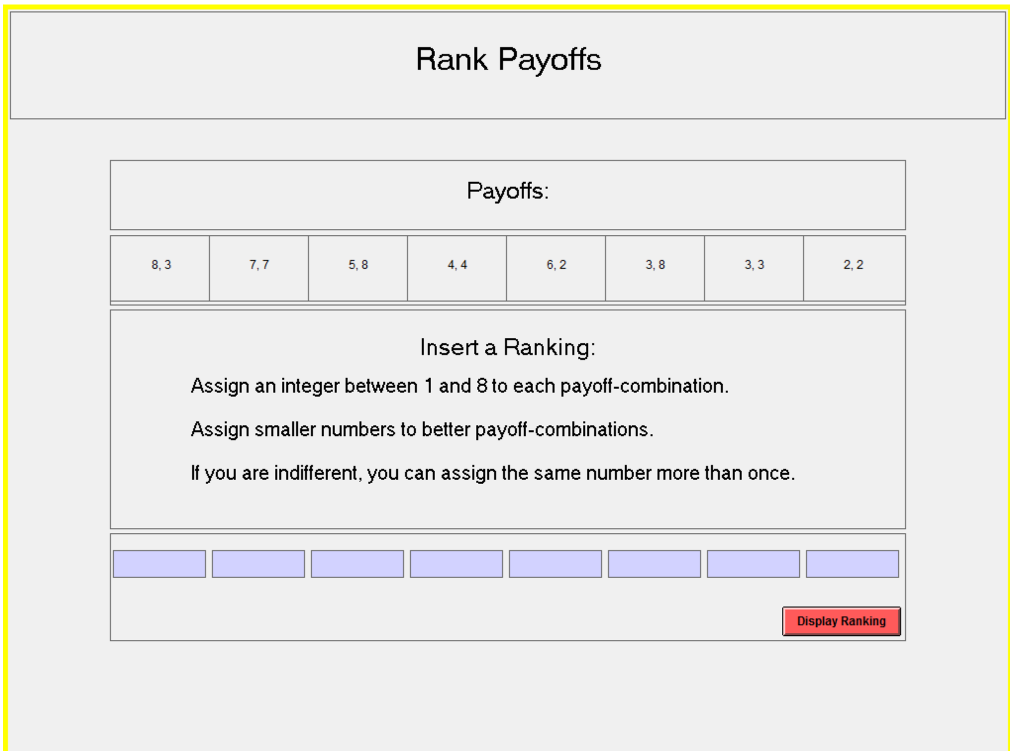
This experiment consists of two parts, which are identical for all participants: In the first part you are shown eight different payoff-combinations, which you are supposed to evaluate. Each of these combinations consists of two numbers (x, y). The first number x corresponds to the amount of Euro that you receive yourself in this situation. The second number y corresponds to the amount that another participant receives. You are supposed to establish a ranking (a so called "preference relationship") over all these payoff-combinations (x, y). That means, you indicate which of these combinations you like best, which one second-best, and so on. The exact procedure will be explained again step by step later on.

The ranking created in this way, as well your decisions in part two of the experiment, will not be revealed to any other participant. After each participant has created such a ranking over the payoff-combinations, part two of the experiment will begin. Both parts of the experiment are run at the computer. Before they start, you are asked several control questions, which shall help you in your understanding of the experiment. For the second part, you will receive separate instructions. At the end of the experiment, there will be a short questionnaire and then you will be paid in cash.

Your total payoff consists of two payments. In order to determine these payments, one of the decisions made in either part 1 or part 2 of the experiment will be randomly selected. Further details will be provided later on.

2 Evaluation of Payoff-combinations

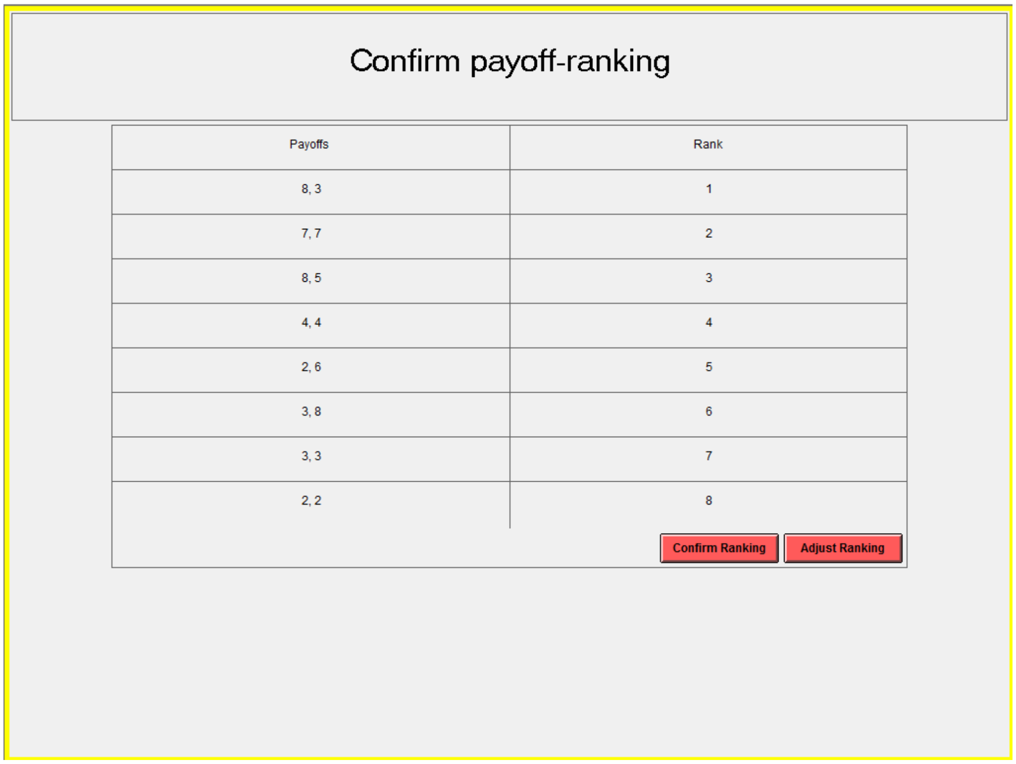
We will now explain the first part of the experiment, the evaluation of payoff-combinations. You will perform this task immediately afterwards at the computer. You will first be shown the following screen:



In the row below "Payoffs" you see the eight different payoff-combinations (x, y), which you are supposed to rank (all amounts are in Euro). The payoff combinations are currently ordered randomly. *(Remember: The left value x is the amount you receive yourself and the right value y is given to a randomly selected other participant.)*

You will now assign a number between 1 and 8 to each of these payoff-combinations. The number 1 corresponds to the first rank, which you shall assign to the combination you like best. Analogously the second rank shall be assigned to your second-best combination and so forth until rank 8, which corresponds to your least preferred combination. If you consider two or more combinations as equally good, you are allowed to assign the same rank/number to them.

After you created your ranking, you will see the following screen:



Here you see the payoff-combinations, ordered according to your previously stated preferences. If you like, you can still make modifications. After all participants confirmed their ranking, the second part of the experiment will begin.

3 Calculation of your Final Payoff

The one and only payoff-relevant decision will be randomly selected at the end of the experiment. Your total payoff depends on whether a decision from the first or the second part of the experiment is selected.

With a probability of $\frac{7}{8}$ a decision of part one will be chosen. In this case, two of the eight payoff-combinations will be randomly selected. The payoff combination that you ranked more highly will then be paid out. (If both combinations have the same rank, one of these two will be randomly selected.). You will receive the first amount, the value x. In addition, every participant receives exactly one additional payment y that corresponds to the second amount y of a payment-combination selected for some other participant.*(The assignment is carried out in such a way that the second amount y from your decision is not distributed to a participant you are interacting with during the experiment or from whom you receive the second amount yourself.)*

Payoff, if selected decision is from part one:

Total payoff = Amount x from own decision + Amount y from decision of some other participant

The probability that a decision from part two is chosen for payment is $\frac{1}{8}$. In that case, payments depend on the actions chosen by the participants in part two. The calculation of the final payoff for this case will be explained in the instructions for this part. *(The random draw will be performed by a participant at the end of the experiment. For that purpose he draws a card from a deck containing 32 cards numbered 1 to 32. The numbers 1-28 correspond to all possible combinations of two out of the eight payoff-pairs (x, y)) from the first part. If a number between 29-32 is drawn, a decision from the second part will be paid out.)*

Treatment Baseline: Instructions Part 2

The second part of the experiment is run at the computer as well. This part consists of four strategic decision situations, in the following referred to as "games". In each of these situations, you will be matched with a different participant as game partner, that means you never interact with the same person twice. You and the other player simultaneously select one of two possible actions. The row player always chooses between one of the two actions "up" and "down" and the column player always decides between the actions "left" and "right". *(For the sake of simplicity, the game will be displayed for every participant in such a way, that he always acts in the role as row player and the game partner in the role as column player.)*

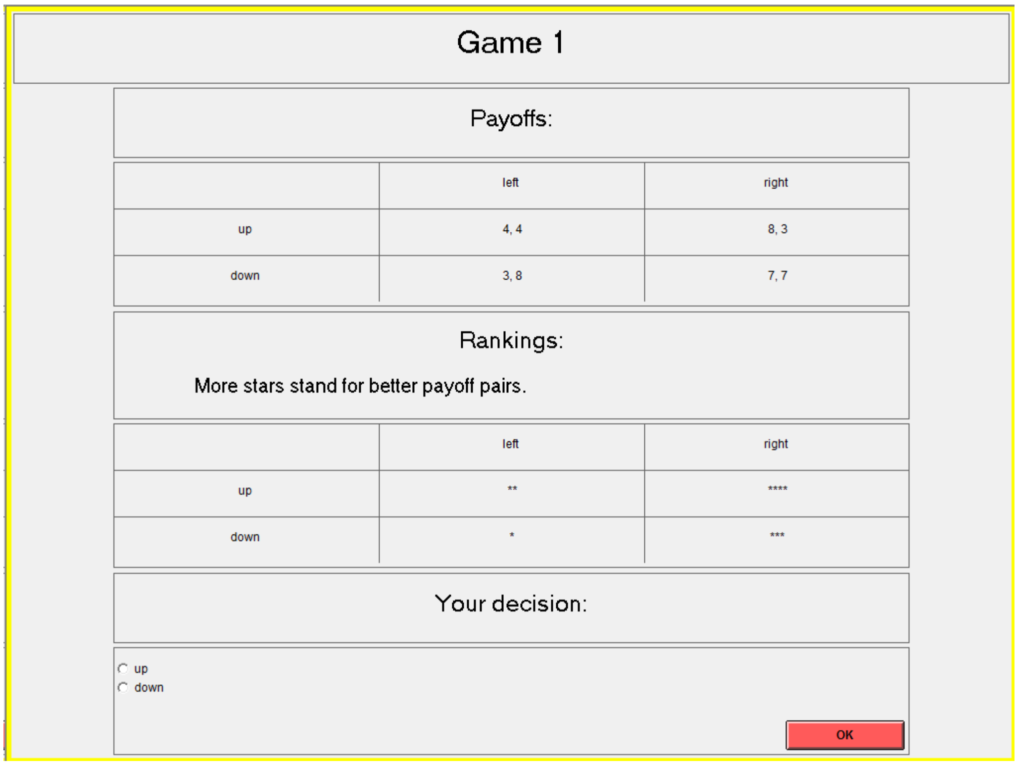
In every game, there are four possible outcomes. Which one of these outcomes is selected depends on the action you chose as well as on the action the other player chooses. The four outcomes are are displayed in the form of a payoff matrix. The combination (x, y) in one cell of the matrix corresponds to the amounts of money the two players receive, if the corresponding actions have been chosen. Analogously to the first part, the left value x indicates the amount of money in Euro that you receive and the right value y corresponds to the payoff of the other player. **The combinations (x, y) are chosen in such a way, that they assume the exact same values as those from the first part of the experiment.** Thus in every game there appear four out of the eight payoff pairs evaluated in part one.

If a situation from the second part is chosen for payment, the involved players receive the payoffs that correspond to the outcome of the game. In contrast to the first part, each player only receives one amount of money from the payoff-relevant decision. In addition, each player is given a fixed payment of 5 Euro.

Total payoff = 5 Euro + Payment x obtained in the selected game

In addition to the monetary payments, you are also shown the ranking of the payoff-pairs used in the current game that you submitted in the first part of the experiment.

In the computer program, you will see the following screen:



For the sake of clarity, not the exact numbers of the ranking will be shown there, but instead 1-4 stars. A value of four stars (****) means that the corresponding payoff-combination was ranked by you as the best combination (among those appearing in the game). Accordingly, the worst combination is marked by one star (*)

Example:
Let us consider the game shown on the screen "Game 1". If, for example, you decide to play "up" and the other player chooses "right", then you receive a payoff of 8 Euros and your game partner a payoff of 3 Euros. Additionally, you can see in the matrix below, that this is your most preferred outcome.

Are there any questions?

If this is not the case, the second part of the experiment will start shortly...

Treatment Info: Instructions Part 1

1. General Information

Welcome to this experiment and thank you very much for your participation! Please switch off your mobile phone now and do not communicate with each other any more. If you have a question, raise your hand, we will come over to your seat and answer it individually. In this experiment, you can earn a substantial amount of money. The amount you earn depends on your own decisions, the decisions of the other participants and on chance. The amount of money earned will be paid out to all participants individually in cash at the end of the experiment.

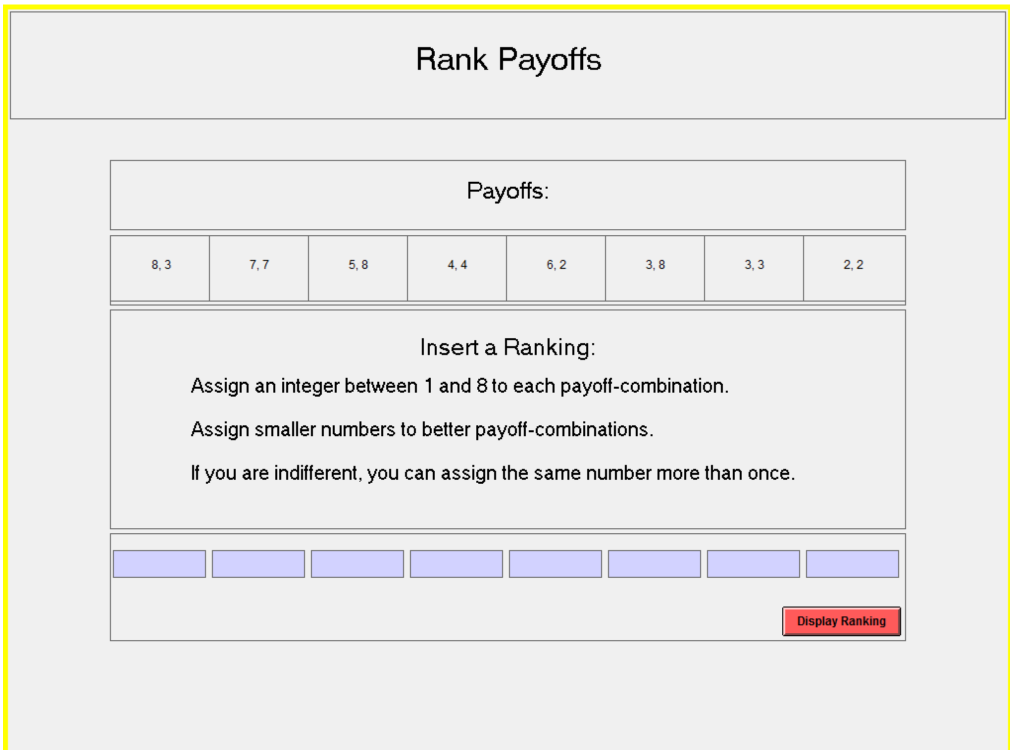
This experiment consists of two parts, which are identical for all participants: In the first part you are shown eight different payoff-combinations, which you are supposed to evaluate. Each of these combinations consists of two numbers (x, y). The first number x corresponds to the amount of Euro that you receive yourself in this situation. The second number y corresponds to the amount that another participant receives. You are supposed to establish a ranking (a so called "preference relationship") over all these payoff-combinations (x, y). That means, you indicate which of these combinations you like best, which one second-best, and so on. The exact procedure will be explained again step by step later on.

After each participant has created such a ranking over the payoff-combinations, part two of the experiment will begin. In this part, the information provided in the first part of the experiment will be used. Two participants at a time will be shown each others' ranking of the payoff-pairs provided in part one of the experiment. In both parts of the experiment, you will interact with other participants using a computer. Before we start, you will be asked several control questions, which shall help you in your understanding of the experiment. For the second part, you will receive separate instructions. At the end of the experiment, there will be a short questionnaire and then you will be paid in cash.

Your total payoff consists of two payments. In order to determine these payments, one of the decisions made in either part 1 or part 2 of the experiment will be randomly selected. Further details will be provided later on.

2. Evaluation of Payoff-combinations

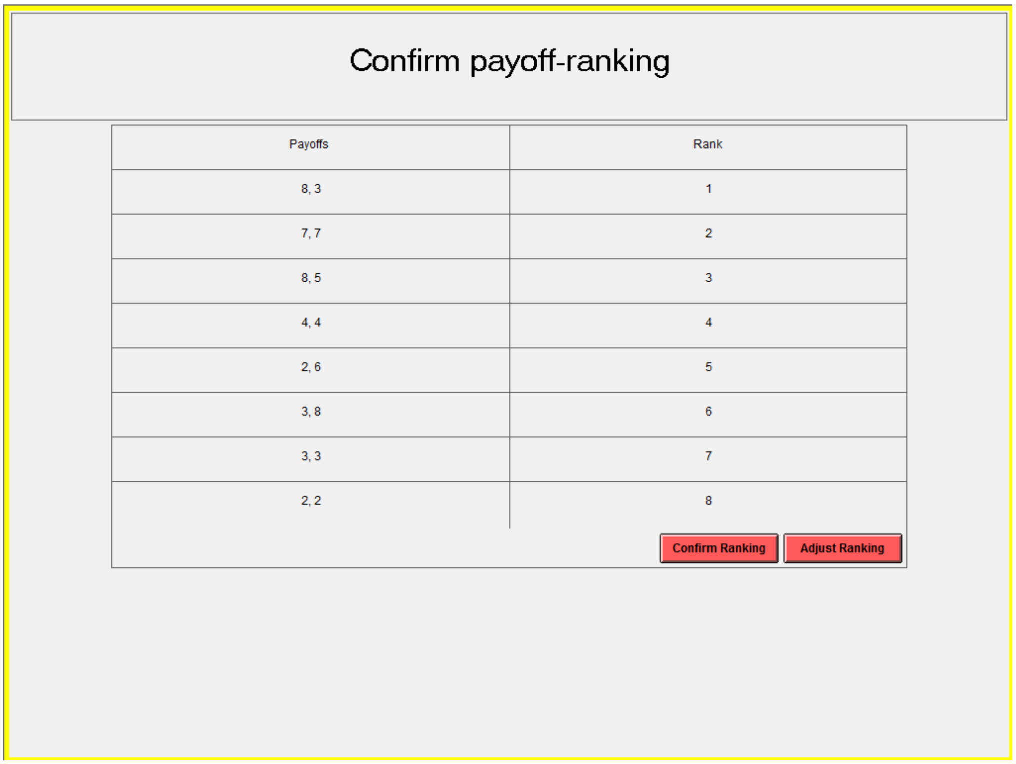
We will now explain the first part of the experiment, the evaluation of payoff-combinations. You will perform this task immediately afterwards at the computer. You will first be shown the following screen:



In the row below "Payoffs" you see the eight different payoff-combinations (x, y), which you are supposed to rank (all amounts are in Euro). The payoff combinations are currently ordered randomly. *(Remember: The left value x is the amount you receive yourself and the right value y is given to a randomly selected other participant.)*

You will now assign a number between 1 and 8 to each of these payoff-combinations. The number 1 corresponds to the first rank, which you shall assign to the combination you like best. Analogously the second rank shall be assigned to your second-best combination and so forth until rank 8, which corresponds to your least preferred combination. If you consider two or more combinations as equally good, you are allowed to assign the same rank/number to them.

After you created your ranking, you will see the following screen:


Confirm payoff-ranking

Payoffs	Rank
8, 3	1
7, 7	2
8, 5	3
4, 4	4
2, 6	5
3, 8	6
3, 3	7
2, 2	8

Here you see the payoff-combinations, ordered according to your previously stated preferences. If you like, you can still make modifications. After all participants confirmed their ranking, the second part of the experiment will begin.

3. Calculation of your Final Payoff

The one and only payoff-relevant decision will be randomly selected at the end of the experiment. Your total payoff depends on whether a decision from the first or the second part of the experiment is selected.

With a probability of $\frac{7}{8}$ a decision of part one will be chosen. In this case, two of the eight payoff-combinations will be randomly selected. The payoff combination that you ranked more highly will then be paid out. (If both combinations have the same rank, one of these two will be randomly selected.). You will receive the first amount, the value x. In addition, every participant receives exactly one additional payment y that corresponds to the second amount y of a payment-combination selected for some other participant.*(The assignment is carried out in such a way that the second amount y from your decision is not distributed to a participant you are interacting with during the experiment or from whom you receive the second amount yourself.)*

Payoff, if selected decision is from part one:

Total payoff = Amount x from own decision + Amount y from decision of some other participant

The probability that a decision from part two is chosen for payment is $\frac{1}{8}$. In that case, payments depend on the actions chosen by the participants in part two. The calculation of the final payoff for this case will be explained in the instructions for this part. *(The random draw will be performed by a participant at the end of the experiment. For that purpose he draws a card from a deck containing 32 cards numbered 1 to 32. The numbers 1-28 correspond to all possible combinations of two out of the eight payoff-pairs (x, y)) from the first part. If a number between 29-32 is drawn, a decision from the second part will be paid out.)*

Treatment Info: Instructions Part 2

The second part of the experiment is run at the computer as well. This part consists of four strategic decision situations, in the following referred to as "games". In each of these situations, you will be matched with a different participant as game partner, that means you never interact with the same person twice. You and the other player simultaneously select one of two possible actions. The row player always chooses between one of the two actions "up" and "down" and the column player always decides between the actions "left" and "right". *(For the sake of simplicity, the game will be displayed for every participant in such a way, that he always acts in the role as row player and the game partner in the role as column player.)*

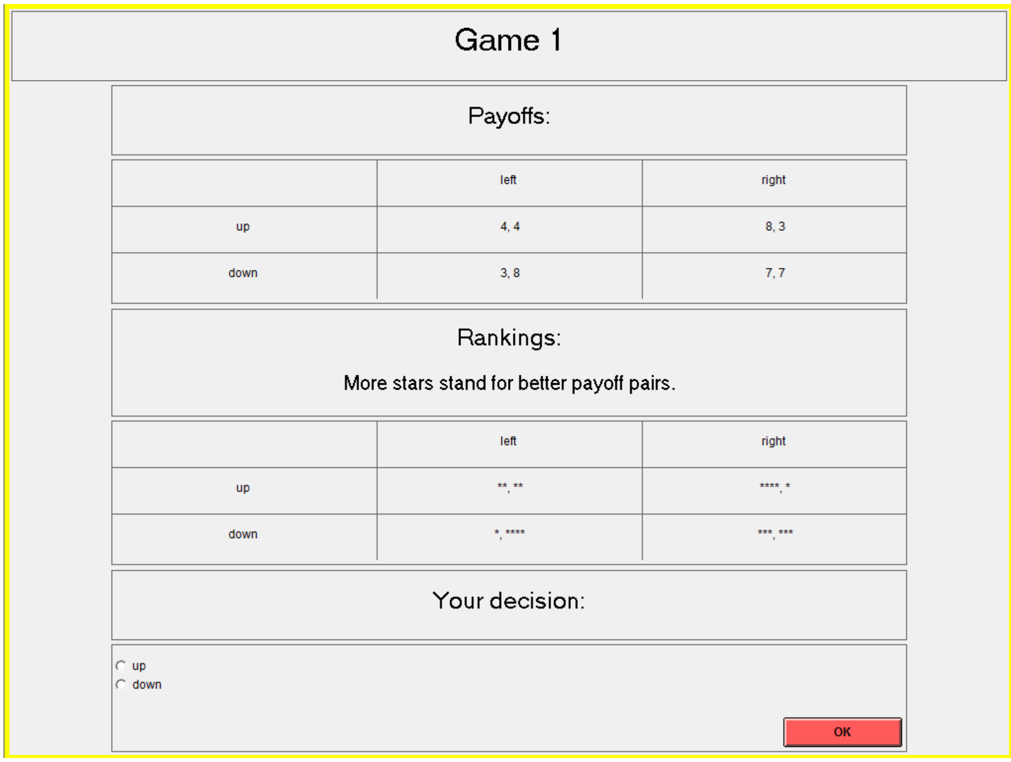
In every game, there are four possible outcomes. Which one of these outcomes is selected depends on the action you chose as well as on the action the other player chooses. The four outcomes are are displayed in the form of a payoff matrix. The combination (x, y) in one cell of the matrix corresponds to the amounts of money the two players receive, if the corresponding actions have been chosen. Analogously to the first part, the left value x indicates the amount of money in Euro that you receive and the right value y corresponds to the payoff of the other player. **The combinations (x, y) are chosen in such a way, that they assume the exact same values as those from the first part of the experiment.** Thus in every game there appear four out of the eight payoff pairs evaluated in part one.

If a situation from the second part is chosen for payment, the involved players receive the payoffs that correspond to the outcome of the game. In contrast to the first part, each player only receives one amount of money from the payoff-relevant decision. In addition, each player is given a fixed payment of 5 Euro.

Total payoff = 5 Euro + Payment x obtained in the selected game

As announced before, you will now receive information about each others' preferences. This means that in addition to the payoff matrix you are shown another matrix below, in which you can see how you and the other player ranked the payoff combinations used in the current game in the first part of the experiment. *Note: you interact with a different partner in every game and therefore the ranking of your opponent may change from one game to another.*

In the computer program, you will see the following screen:



For the sake of clarity, not the exact numbers of the ranking will be shown there, but instead 1-4 stars. A value of four stars (****) means that the corresponding payoff-combination was ranked by you as the best combination (among those appearing in the game). Accordingly, the worst combination is marked by one star (*)

Example:
Let us consider the game shown on the screen "Game 1". If, for example, you decide to play "up" and the other player chooses "right", then you receive a payoff of 8 Euros and your game partner a payoff of 3 Euros. Additionally, you can see in the matrix below, that this is your most preferred outcome, but the least preferred outcome of the other player.

Are there any questions?

If this is not the case, the second part of the experiment will start shortly...